Does Teaching Experience Help? Differences in the Assessment of Tutees' Understanding Between Teacher Tutors and Student Tutors

Stephanie Herppich (herppich@ipn.uni-kiel.de)
Institute for Science and Mathematics Education at the University of Kiel
Olshausenstrasse 62, 24098 Kiel, Germany

Jörg Wittwer (joerg.wittwer@sowi.uni-goettingen.de)
University of Göttingen, Educational Institute
Waldweg 26, 37073 Göttingen, Germany

Matthias Nückles (matthias.nueckles@ezw.uni-freiburg.de)
University of Freiburg, Department of Educational Science, Instructional and School Research
Rempartstrasse 11, 79098 Freiburg, Germany

Alexander Renkl (renkl@psychologie.uni-freiburg.de)
University of Freiburg, Department of Psychology, Developmental and Educational Psychology
Engelbergerstrasse 41, 79085 Freiburg, Germany

Abstract

Tutors often have difficulty in accurately assessing their tutees' understanding. However, it is unclear how tutors' professional experience influences their assessment accuracy. Therefore, we conducted a study with $N = 46$ tutor-tutee dyads and compared the accuracy with which teacher tutors and student tutors assessed a tutee's understanding of the human circulatory system at the level of mental models. Results showed that both groups of tutors faced similar difficulties in assessing a tutee's understanding. However, whereas teacher tutors' assessment accuracy remained constant in the course of tutoring, student tutors' assessment accuracy decreased. Moreover, teacher tutors more accurately self-assessed their assessment accuracy than student tutors. Although teacher tutors process diagnostic information more accurately than student tutors, both groups of tutors seem to be overwhelmed by processing all information making up a tutee's mental model. Hence, regardless of their professional experience, tutors need to be supported in assessing a tutee's understanding.

Keywords: assessment accuracy, human tutoring, teacher tutors, student tutors

Introduction

To be effective, instruction should be adapted to the learner (Kalyuga, 2007). Human one-to-one tutoring is a method that offers many opportunities to adapt instruction to a learner's understanding (Snow & Swanson, 1992). However, adaptive instruction makes it necessary for tutors to assess a tutee's understanding accurately. Research has shown that tutors often have difficulty with collecting diagnostically relevant information about a tutee. This seems to be true irrespective of whether experienced classroom teachers (i.e., teacher tutors) or university students (i.e., student tutors) act as tutors.

For example, Chi, Roy, and Hausmann (2008) found that a teacher tutor was not responsive to a tutee's level of understanding and comprehension problems (see also Putnam, 1987). Similarly, Chi, Siler, and Jeong (2004) observed that student tutors overestimated a tutee's correct understanding and underestimated a tutee's incorrect understanding. Chi et al. (2004) explained this finding with the tutors' bias to use their own normative understanding as a basis for assessing a tutee's understanding (see also Graesser, Person, & Magliano, 1995).

However, none of these studies compared the assessment skills of teacher tutors with the assessment skills of student tutors empirically although there is much theoretical research on such differences between expert tutors and novice tutors (e.g., Graesser, D'Mello, & Cade, 2009). Therefore, it remains open as to what extent the assessment accuracy of teacher tutors differs from the assessment accuracy of student tutors. In this article, we present an empirical study to shed light on this question. To elucidate possible differences between teacher tutors and student tutors, we draw on findings from research on judgments about learners outside the tutoring context. We integrate these findings by applying Nickerson's (1999) theory on the development of a model of another person's knowledge.

Outside the Tutoring Context: Accuracy of Judgments About Learners

Research has shown that classroom teachers are relatively accurate in knowing how a learner performs relative to other learners in a class (Hoge & Coladarci, 1989). However, when looking at the absolute level of the classroom teachers' estimates of a learner's performance, classroom teachers often overestimate a learner's performance (e.g., Bates & Nettelbeck, 2001see). Hence, classroom teachers seem to have difficulty with assessing a learner's individual understanding.

Nevertheless, research also suggests that classroom teachers and university students differ in their ability to accurately assess a learner's performance. For example,

Dünnebier, Gräsel, and Krolak-Schwerdt (2009) showed that classroom teachers, on the one hand, could quite accurately grade the performance of a fictitious learner in a German test. They were not strongly influenced by a grade believed to be provided by an experienced colleague. This was especially true if classroom teachers were made to believe that they had to give an important educational recommendation. University students, on the other hand, were strongly influenced by the grade of the 'experienced colleague'. They generally used this grade as an anchor (cf. Tversky & Kahneman, 1974) for their judgment. Similarly, Krolak-Schwerdt, Böhmer, and Gräsel (2009) found that classroom teachers flexibly changed between different modes of processing when assessing a fictitious learner's performance. The type of processing they showed depended on the aim associated with the assessment procedure. That is, when the classroom teachers believed that they merely had to form an impression of the fictitious learner they paid most attention to stereotypical information about the learner. Conversely, when the classroom teachers believed that they had to give an important educational recommendation they paid most attention to individual information about the learner. University students, however, failed to display such different modes of processing. Krolak-Schwerdt et al. (2009) attributed these differences between teachers and university students to the fact that teachers possess more knowledge about learners than do university students. Finally, Südkamp and Möller (2009) examined whether university students would be able to use information about fictitious learners in the course of teaching to improve their accuracy in assessing the learners' performance. They showed that the university students generally overestimated the learners' performance and did not become more accurate in their assessment.

To summarize, the results illustrate that classroom teachers are quite accurate at assessing learners under some conditions. The university students' assessment accuracy, however, seems to be more limited. Moreover, university students appear to be less flexible in processing information about a learner than classroom teachers.

A Theoretical Framework for Understanding Differences in the Assessment Between Teacher Tutors and Student Tutors

To understand the differences in the assessment performance between classroom teachers and university students, we draw on Nickerson's theory (1999), which describes how people in general construct a model of another person's knowledge. In the first phase, one's own knowledge serves as an anchor for building a default model of another person's knowledge. In the second phase, the default model is transformed into a more person-specific model. This is done by deriving information, amongst others, about the community to which the person belongs. In the third phase, a more individualized model is modified on an ongoing basis in accordance with information obtained when interacting with the person.

Classroom teachers as well as university students enter a teaching situation equipped with their own person specific knowledge, which they probably use to build a default model of a learner's knowledge (cf. Nickerson, 1999; phase 1). Yet, classroom teachers who are experienced teachers dedicate a lot of time to engaging in assessment activities (e.g., Martínez, Stecher, & Borko, 2009) and have usually gained a differentiated categorical knowledge about learners, their knowledge, and how they learn (Krolak-Schwerdt et al., 2009). This knowledge should help them to adjust their model about a learner's knowledge in the tutoring situation (cf. Nickerson, 1999; phase 2). Hence, their assessments of a learner's understanding should be more accurate than those of university students who do not have such a teaching experience.

Classroom teachers should also have developed routines to face the multiple demands of teaching. Compared to university students, they should be less likely affected by a cognitive overload that occurs when teaching demands exceed the limited capacity of the working memory (Feldon, 2007). Consequently, classroom teachers should be able to spend more cognitive resources on flexibly adapting their model of a learner's knowledge and become more accurate at assessing a learner's understanding in the course of teaching (cf. Nickerson, 1999; phase 3).

Moreover, the experience of classroom teachers should positively influence the assessment of their own assessment skills. This is because, as Dunning, Johnson, Ehrlinger, and Kruger (2003) argue, the skills that are necessary to self-assess one's own task performance are strongly associated to those skills that are necessary to accomplish the task. Hence, classroom teachers are probably more self-aware of their assessment skills than university students.

Hypotheses

We present an empirical study in which we examined the extent to which the assessment accuracy of teacher tutors differs from the assessment accuracy of student tutors. More specifically, we addressed the following hypotheses:

1) Teacher tutors are more accurate at assessing a tutee's mental model than student tutors.
2) The accuracy of teacher tutors' assessments increases more strongly in the course of tutoring than the accuracy of student tutors' assessments.
3) Teacher tutors more accurately self-assess their assessment skills than student tutors.

Method

Sample and Design

A total of $N = 46$ dyads of tutors and tutees participated in the empirical study. Twenty-one tutors were biology teachers (teacher tutors). The mean age of the teacher tutors was $M = 44.05$ years ($SD = 11.76$). Of the teacher tutors, 11 tutors were female and 10 tutors were male. On average, the teacher tutors had $M = 13$ years ($SD = 12.30$) of professional experience as a biology teacher. Twenty-five tutors were

university students of biology (student tutors). The mean age of the student tutors was $M = 22.24$ years ($SD = 2.83$). Of the student tutors, 21 tutors were female and 4 tutors were male. A multiple-choice test showed that the teacher tutors ($M = 12.43$, $SD = 3.43$) and the student tutors ($M = 11.56$, $SD = 3.86$) had comparable knowledge about the human circulatory system, $F(1, 44) = 0.63$, $p = .43$, $\eta^2 = .01$ (small effect). Tutees were grade 7 students from Realschulen (i.e., schools from the middle track of the German school system). Of the tutees, 19 were female and 27 were male. Their mean age was $M = 12.65$ years ($SD = 0.53$).

The main dependent variable in this study was the accuracy with which the tutors assessed a tutee's understanding of the human circulatory system at the level of mental models. In addition, we examined the extent to which the tutors accurately self-assessed their assessment accuracy.

Materials

Textbook (Tutee and Tutor) In the tutoring session, the tutor and the tutee engaged in a dialogue on the basis of a passage about the human circulatory system, which was previously used by Chi, Siler, Jeong, Yamauchi, and Hausmann (2001). We adapted this passage for the present study by deleting and reformulating sentences. Each of the remaining 59 sentences of the passage was printed on a separate sheet of paper. The sentences were presented to the tutor and the tutee in a ring binder.

Drawings of the Human Circulatory System (Tutee and Tutor) On a sheet of paper, the outline of a human body was displayed. The tutees were asked to draw the blood path of the circulatory system into the human body and to explain the blood path. The explanations were audiotaped. By using this methodology, which was originally developed by Chi et al. (2004), we assessed a tutee's conceptual understanding about the human circulatory system at the level of mental models.

To code the tutees' and the tutors' drawings and explanations of the human circulatory system, we adapted a classification scheme originally developed by Azevedo, Cromley, and Seibert (2004). On the basis of this classification scheme, the drawings were assigned a score between 0 and 11. The scores reflect distinguishable types of correct and incorrect mental models with scores from 0 to 9 indicating different types of incorrect mental models and with scores from 10 to 11 indicating a correct mental model.

Self-Assessment of Assessment Accuracy (Tutor) At the end of tutoring, we asked the tutors to self-assess the accuracy with which they had assessed a tutee's understanding at the level of mental models in the midst and at the end of tutoring. To do so, the tutors used a 4-point rating scale ranging from 1 (= *very inaccurate*) to 4 (= *very accurate*).

Procedure

Each tutoring session was divided into three phases: pre-test phase, tutoring phase, and post-test phase. It lasted about 3 hours.

Pre-Test Phase In the pre-test phase, the tutees were asked to draw the blood path of the human circulatory system in the outline of a human body and to explain the blood path as they knew it. Afterwards, both the tutors and the tutees individually read the passage about the human circulatory system.

Tutoring Phase The dyads of tutors and tutees read each sentence of the passage about the human circulatory system and engaged in a dialogue about each sentence. After the 33rd sentence, tutoring was interrupted and the dyads were separated. The tutees were asked to draw and explain the blood path of the human circulatory system. To measure what the tutors thought that the tutees would know about the blood path, the tutors were required to draw and explain the tutees' mental model of the human circulatory system. After accomplishing this task, tutoring was continued.

Post-Test Phase After completing the tutorial dialogue, the dyads of tutors and tutees were separated again and asked to draw and explain the blood path of the human circulatory system. Afterwards, the tutors completed the self-assessment rating scale.

Results

Tutors' Assessment Accuracy

In a first step, we analyzed the extent to which teacher tutors and student tutors accurately assessed the level of correctness of the tutees' mental model of the human circulatory system in the midst of tutoring and at the end of tutoring. To do so, we compared the score assigned to the tutees' actual mental model of the human circulatory system with the score assigned to the tutors' assumed mental model of the human circulatory system. The scores were subjected to a repeated-measures ANOVA with type of tutor as between-subjects factor.

Assessment Accuracy in the Midst of Tutoring The results showed that the tutors overall significantly overestimated the level of correctness of the tutees' mental model of the human circulatory system in the midst of tutoring ($M_{\text{Tutee}} = 6.52$, $SD_{\text{Tutee}} = 2.65$; $M_{\text{Tutor}} = 8.09$, $SD_{\text{Tutor}} = 2.48$), $F(1, 44) = 13.50$, $p < .001$, $\eta^2 = .24$ (large effect). The teacher tutors ($M_{\text{Tutee}} = 6.10$, $SD_{\text{Tutee}} = 2.74$; $M_{\text{Tutor}} = 8.29$, $SD_{\text{Tutor}} = 2.59$) were not significantly more accurate in assessing the level of correctness of the tutees' mental model of the human circulatory system than the student tutors ($M_{\text{Tutee}} = 6.88$, $SD_{\text{Tutee}} = 2.57$; $M_{\text{Tutor}} = 7.92$, $SD_{\text{Tutor}} = 2.41$), $F(1, 44) = 1.71$, $p = .20$, $\eta^2 = .04$ (small effect).

Assessment Accuracy at the End of Tutoring The results showed that the tutors overall significantly overestimated the level of correctness of the tutees' mental model of the human circulatory system at the end of tutoring (M_{Tutee} = 7.93, SD_{Tutee} = 2.69; M_{Tutor} = 10.07, SD_{Tutor} = 0.88), $F(1, 44) = 27.94$, $p < .001$, $\eta^2 = .39$ (large effect). The teacher tutors (M_{Tutee} = 8.05, SD_{Tutee} = 2.67; M_{Tutor} = 10.00, SD_{Tutor} = 0.84) were not significantly more accurate in assessing the level of correctness of the tutees' mental model of the human circulatory system than the student tutors (M_{Tutee} = 7.84, SD_{Tutee} = 2.75; M_{Tutor} = 10.12, M_{Tutor} = 0.93), $F(1, 44) = 0.17$, $p = .68$, $\eta^2 = .01$ (small effect).

Changes in the Assessment Accuracy in the Course of Tutoring

In a second step, we examined whether tutors would become more accurate in assessing the tutees' understanding in the course of tutoring. To do so, we subtracted the score assigned to the tutees' actual mental model of the human circulatory system from the score assigned to the tutors' assumed mental model of the human circulatory system and compared the difference scores obtained in the midst of tutoring with the difference scores obtained at the end of tutoring. The difference scores were subjected to a repeated-measures ANOVA with the type of tutor as between-subjects factor.

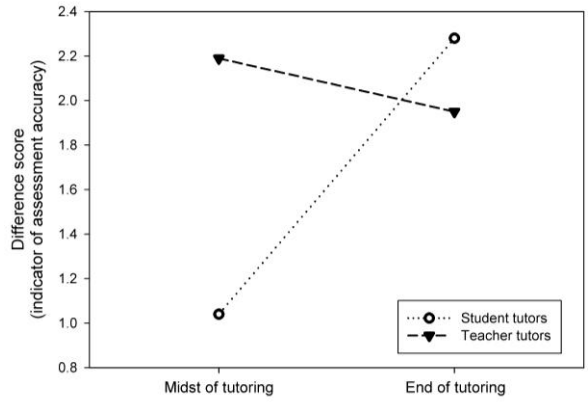

Figure 1: Interaction effect between measurement point and type of tutor on assessment accuracy.

On average, the tutors did not improve their assessment accuracy in the course of tutoring (M_{Midst} = 1.57, SD_{Midst} = 2.99; M_{End} = 2.13, SD_{End} = 2.68), $F(1, 44) = 2.08$, $p = .16$, $\eta^2 = .05$ (medium effect). However, the results indicated a significant interaction effect $F(1, 44) = 4.53$, $p = .04$, $\eta^2 = .09$ (medium effect). Whereas the student tutors became more inaccurate in assessing the tutees' mental model from the midst to the end of tutoring (M_{Midst} = 1.04, SD_{Midst} = 3.21; M_{End} = 2.28, SD_{End} = 2.76) the teacher tutors became slightly more accurate in assessing the tutees' mental model from the midst to the end of tutoring (M_{Midst} = 2.19, SD_{Midst} = 2.66; M_{End} = 1.95, SD_{End} = 2.64; cf. Figure 1).

Self-Assessment of the Accuracy in Assessing a Tutee's Understanding of the Human Circulatory System

In a third step, we examined the extent to which the tutors accurately self-assessed the accuracy with which they assessed the tutees' understanding of the human circulatory system. First, we performed a repeated-measures ANOVA with the tutors' self-assessments in the midst of tutoring and at the end of tutoring as repeated measures and the type of tutor as between-subjects factor. The results showed that the tutors assumed their assessment accuracy to be significantly higher at the end of tutoring ($M = 2.93$, $SD = 0.54$) than in the midst of tutoring ($M = 2.33$, $SD = 0.60$), $F(1, 43) = 30.13$, $p < .001$, $\eta^2 = .41$ (large effect). In addition, the student tutors ($M = 2.76$, $SD = 0.39$) self-assessed their assessment accuracy as being significantly higher than the teacher tutors ($M = 2.48$, $SD = 0.44$), $F(1, 43) = 5.32$, $p = .03$, $\eta^2 = .11$ (medium effect).

Second, in order to analyze the extent to which the tutors' self-assessments reflected their assessment accuracy, we computed correlations between the tutors' self-assessments and their assessment accuracy as reflected by the difference between the score assigned to the tutors' assumed mental model of the human circulatory system and the score assigned to the tutees' actual mental model of the human circulatory system. The correlations for the self-assessments of the teacher tutors were significant (r_{Midst} = -.47, p_{Midst} = .04; r_{End} = -.56, p_{End} = .01). Hence, the more the teacher tutors assumed their assessment to be accurate, the more accurate their assessment actually was, as indicated by a lower difference between the tutors' assumed mental model and the tutees' actual mental model. In contrast, the correlations for the self-assessments of the student tutors were not significant (r_{Midst} = .14, p_{Midst} = .49; r_{End} = .00, p_{End} = .99).

Discussion

This study compared the accuracy with which teacher tutors and student tutors assessed a tutee's understanding of the human circulatory system. We found that the tutors overall overestimated the level of correctness of the tutees' mental model of the human circulatory system in the midst and at the end of the tutoring session. In contrast to our expectations, teacher tutors were not more accurate at assessing the tutees' understanding than student tutors.

However, there were more subtle differences between teacher tutors and student tutors. First, student tutors became more inaccurate in their assessments in the course of tutoring. As a result, their overestimations of the tutees' understanding were even more pronounced at the end of tutoring than in the midst of tutoring. Conversely, teacher tutors became slightly, albeit insignificantly, more accurate. Overall, tutors actually did not become more accurate in assessing the tutees' understanding. Yet, they assumed their

assessment accuracy to be higher at the end of tutoring than in the midst of tutoring. With regard to their self-assessments of assessment accuracy, the student tutors had, second, the impression of being even more accurate than the teacher tutors. However, our analyses revealed no significant differences in the assessment performance between teacher tutors and student tutors. Third, the results showed that the teacher tutors were fairly accurate in knowing the extent to which they accurately assessed a tutee's understanding of the human circulatory system whereas the student tutors failed to do so.

Our result that tutors generally overestimated the correctness of the tutees' understanding is in line with previous research (Chi et al., 2004). Following Chi et al. (2004) and Nickerson's (1999) theory on the development of a model of another person's knowledge, it can be assumed that the tutors might have tended to rely too heavily on their own correct and readily available understanding as an anchor to construct a mental model of the tutees' understanding. As a result, the tutors failed to adjust this anchor sufficiently in order to account for differences between their own and the tutees' understanding (see also Wittwer, Nückles, Landmann, & Renkl, 2010).

Moreover, teacher tutors were not more accurate in assessing a tutee's understanding than student tutors. We propose two possible explanations for this finding. First, the task of drawing and explaining a tutee's mental model of the human circulatory system might have been relatively unfamiliar to all tutors in this study. Second, assessing a tutee's understanding at the level of mental models is rather difficult because all pieces of information that make up a tutee's mental model must be retrieved by the tutors from their episodic memory and integrated into the current mental model in working memory. Hence, this task might have been too difficult for student tutors and teacher tutors to produce differences in their assessment accuracy. Indeed, further results of the study not reported here showed that a task that was more familiar and less resource-demanding produced differences in the assessment accuracy in favor of teacher tutors (cf. Wittwer, Herppich, Nückles, & Renkl, submitted).

Apart from this, teacher tutors seemed to be aware of the tutees' actual understanding at least to some degree because they slightly improved their assessment accuracy during tutoring. Being experienced teachers, they probably could draw on their differentiated knowledge about learners and their understanding (Krolak-Schwerdt et al., 2009). Although differences between the teacher tutors' assessment accuracy and the student tutors' assessment accuracy did not become apparent, the teacher tutors' model of a tutee's understanding might have been more elaborate. Moreover, teacher tutors probably have certain teaching routines at their disposal that help them cope with the multiple cognitive demands of tutoring. Thus, they might have been able to spend at least some cognitive resources on flexibly adapting their assessment to the tutees' understanding in the course of tutoring and were less influenced by their own understanding (cf. Dünnebier et al., 2009). Student tutors, on the contrary, lack this teaching experience (and knowledge about tutees) and might have fully relied on their own correct understanding as a basis for their assessment. In addition, they might "by default" have assumed that the tutees' understanding would be most similar to their own understanding at the end of tutoring because then all (normatively correct) contents of the textbook passage had been discussed and possibly learned by the tutees.

The difference between teacher tutors and student tutors with regard to changes in their assessment accuracy are reflected in the differences between teacher tutors' and student tutors' self-assessments of assessment accuracy. On the one hand, the student tutors particularly overestimated their assessment accuracy. On the other hand, the teacher tutors rather accurately self-assessed their assessment accuracy. Overall, these findings are in line with the considerations of Dunning et al. (2003). They report that people who are not proficient in accomplishing a certain task usually overestimate their performance in the task. People who are proficient in accomplishing a task, however, usually show rather accurate self-assessments. As Dunning et al. (2003) argue, this is because the skills that are necessary to self-assess one's own performance and those skills that are necessary to accomplish a task are connected by meta-cognitive monitoring and evaluation. It can be assumed that differences in cognitive processing between teacher tutors as experienced teachers and student tutors are reflected by differences in metacognitive skills (e.g., Borko & Livingston, 1989. Teacher tutors have probably developed those meta-cognitive skills that enable them to accurately self-assess their assessment skills. Student tutors, on the contrary, might miss such meta-cognitive skills.

What are the implications of our study and what are the directions for future research? First, we found differences between teacher tutors and student tutors in assessment-related variables (i.e., change in assessment accuracy, self-assessment of assessment accuracy) that point to differences in the processing of diagnostically relevant information about a tutee's understanding between teacher tutors and student tutors. More direct evidence, however, is still needed. Therefore, it would be interesting to analyze the actual tutoring interactions to see if differences between teacher tutors and student tutors materialize in teaching and in "online" assessment behavior. Second, it would be interesting to examine which characteristics of tutors influence their assessment (accuracy) of a tutee's understanding. In this regard, it would also be interesting to see whether teacher tutors and student tutors differ in these characteristics and whether these possible differences in characteristics are associated with differences in their assessment (accuracy). A detailed knowledge of the characteristics that influence assessment accuracy would help to comprehend why certain tutors are more accurate at assessing than other tutors. Third, the assessment accuracy of all tutors in our study was suboptimal. This might strongly limit the tutors' possibilities to adapt their

instruction to the learner. Hence, the effectiveness of tutoring and teaching in general would probably benefit from trainings that aim to improve the assessment skills of tutors and teachers. Such trainings could be informed by an analysis of teaching actions in the actual tutoring interactions. Fourth, to examine the ecological validity of our findings future studies should test whether the results can be replicated in more naturalistic tutoring situations.

Acknowledgments

We would like to thank our research assistants Julian Etzel, Imme Husmeier, Tatjana Scharping, Anika Schoneville, and Raoul Zimmermann for their help with many practical aspects of the project. Moreover, we would like to thank Prof. Dr. Ute Harms and Prof. Dr. Helmut Prechtl for their expert advice with regard to biological subject matters. This research was supported by grants from the German Science Foundation DFG (WI 3348/2-1).

References

Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology, 29*, 344-370.

Bates, C., & Nettelbeck, T. (2001). Primary school teachers' judgments of reading achievement. *Educational Psychology, 21*, 177-187.

Borko, H. & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal, 26*, 473–498.

Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning, *Cognitive Science, 32*, 301-341.

Chi, M. T. H., Siler, S., & Jeong, H. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction, 22*, 363-387.

Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471-533.

Cromley, J. G., & Azevedo, R. (2005). What do reading tutors do? A naturalistic study of more and less experienced tutors in reading. *Discourse Processes, 40*, 83-113.

Dünnebier, K., Gräsel, C., & Krolak-Schwerdt, S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung. Eine experimentelle Studie zu Ankereffekten. [Biases in teachers' assessments of student performance: An experimental study of anchoring effects.]. *Zeitschrift für Pädagogische Psychologie, 23*, 187-195.

Dunning, D., Johnson, K., Ehrlinger, J., and Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*, 83–87.

Feldon, D. F. (2007). Cognitive load in the classroom: The double-edged sword of automaticity. *Educational Psychologist, 42*, 123-137.

Graesser, A. C., D'Mello, S., & Cade, W. (2009). Instruction based on tutoring. In R. E. Mayer and P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 408-426). New York: Routledge.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology, 9*, 495-522.

Hoge, R., & Coladarci, T. (1989). Teacher-based judgments of academic achievement. *Review of Educational Research, 59*, 297-313.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509-539.

Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess. Der Lehrer als flexibler Denker. [Goal-directed processing of students' attributes: The teacher as "flexible thinker".]. *Zeitschrift für Pädagogische Psychologie, 23*, 175-186.

Martínez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment, 14*, 78-102.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737-759.

Putnam, R. T. (1987). Structuring and adjusting content for students: A study of live and simulated students tutoring of addition. *American Educational Research Journal, 24*, 13-48.

Snow, R. E., & Swanson, J. (1992). Instructional psychology: Aptitude, adaptation, and assessment. *Annual Review of Psychology, 43*, 583-626.

Südkamp, A., & Möller, J. (2009). Referenzgruppeneffekte im Simulierten Klassenraum: Direkte und indirekte Einschätzungen von Schülerleistungen. [Reference-group-effects in a simulated classroom: Direct and indirect judgments.]. *Zeitschrift für Pädagogische Psychologie, 23*, 161-174.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Wittwer, J., Herppich, S., Nückles, M., & Renkl, A. (2011). *Does it Make a Difference? Investigating the Assessment Accuracy of Teacher Tutors and Student Tutors.* Manuscript submitted for publication.

Wittwer, J., Nückles, M., Landmann, N., & Renkl, A. (2010). Can tutors be supported in giving effective explanations? *Journal of Educational Psychology, 102*, 74-89.